

Getting Started with Data Management & DMPTool at WHOI

Audrey Mickle

MBLWHOI Library
January 26th, 2017



Overview

1. Why is data management important?
2. What is a data management plan?
3. The DMPTool

Overview

1. **Why is data management important?**
2. What is a data management plan?
3. The DMPTool

Why is Data Management Important?

Data sharing allows for **reproducibility**, **transparency**, and **data reuse** in research.

Sharing is easier if **data are managed well** from the start of a project.

Data Management & Sharing Mandates

- Journals – PLOS, Nature, JDAP partners
- Funders – NSF, NIH...
- Office of Science & Technology Policy mandate, February 2013



Overview

1. Why is data management important?
2. **What is a data management plan?**
3. The DMPTool

What is a data management plan?

A document that describes what you
will do with your data
during your research
and after you complete your project

Why prepare a DMP?

- Saves time
- Increases research efficiency
- Satisfies funder requirements
- Makes reproducibility & sharing easier

A DMP is a Living Document

- Keep your plan current
- Incorporate changes
- Use as a guide for daily activities

Where to Start?

Small & Simple

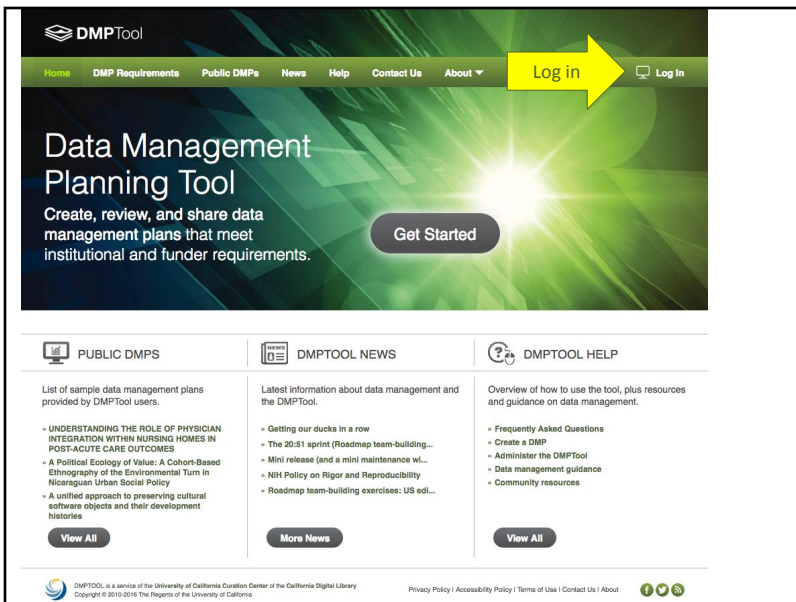
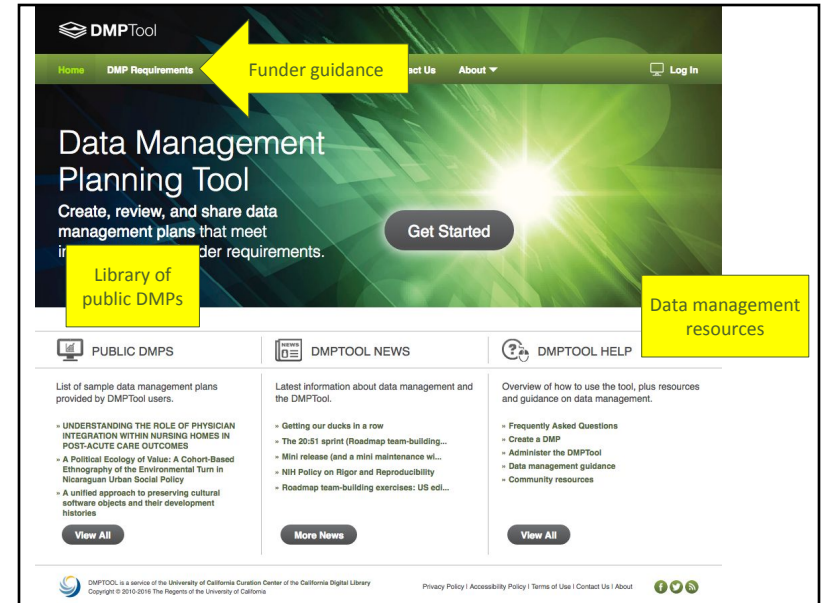
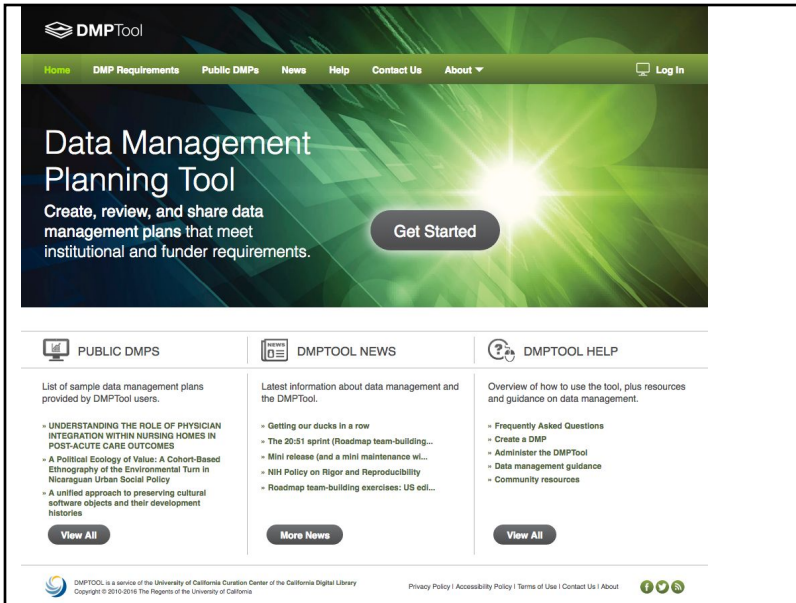
- Document what you know **now**
- Share the plan with your team
- Avoid procrastination and immobilization

Overview

1. Why is data management important?
2. What is a data management plan?
3. **The DMPTool**

DMPTool.org

goo.gl/SLi2jP



DMPTool Welcome! Stephanie Simms

Home My Dashboard DMP Requirements Public DMPs News Help Contact Us About Log Out

Woods Hole Oceanographic Institution

UC3 Contact UC3

My Dashboard | My DMPs | Create New DMP | My Profile

MY PROFILE

Personal Information Notification Preferences

Please update your account information.

Name

* Username Stephanie.Simms-ucoop.edu@uuc


* First name Stephanie

* Last name Simms

Contact Information

* Email Stephanie.Simms@ucoop.edu

Institution University of California, Office of the President

ORCID  orcid.org/0000-0001-9129-3790 Edit Delete

API Token

My Token 63aacc80425b8d3e5849f1f870cc266 Generate New Token Remove Token

Save Cancel

Set up your profile & notification preferences

Add your ORCID

DMPTool Welcome! Stephanie Simms

Home My Dashboard DMP Requirements Public DMPs News Help Contact Us About Log Out

Woods Hole Oceanographic Institution

UC3 Contact UC3

My Dashboard | My DMPs | Create New DMP | My Profile

OVERVIEW

My DMPs My DMPs under Review (where applicable)

8 plans I own No DMPs are under review.

0 plans I co-own

Create New DMP

Dashboard overview of DMPs

Create new DMP

DMPTool is a service of the University of California Curation Center of the California Digital Library
Copyright © 2010-2016 The Regents of the University of California

Privacy Policy | Accessibility Policy | Terms of Use | Contact Us | About

DMPTool Welcome! Stephanie Simms

Home My Dashboard DMP Requirements Public DMPs News Help Contact Us About Log Out

Woods Hole Oceanographic Institution

UC3 Contact UC3

My Dashboard | My DMPs | Create New DMP | My Profile

CREATE NEW DMP

Start with a DMP Template

To create a new DMP, select a function from the list below.

Select Template >>

Either start a new plan

Copy an Existing DMP

The existing DMPs in this list are either publicly shared by any user, shared within your institution by other DMP creators, or are plans that you have previously created.

Select an existing DMP below and add text to the template.

Name	Owner
<input type="radio"/> USGS CDRECV DMP	Calli Jenkinson
<input type="radio"/> A unified approach to preserving cultural software objects and their development histories	DMP dmpcurator
<input type="radio"/> Atmospheric CO2 Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013	DMP dmpcurator
<input type="radio"/> Daymet Follow-On: Surface Weather Data with Uncertainty Quantification for Terrestrial Ecosystem Process Models	DMP dmpcurator
<input checked="" type="radio"/> Multimedia Text Annotation for Students	DMP dmpcurator

1 2 3 4 5 ... Next > Last >

Copy this DMP >>

Or copy an existing DMP

DMPTool Welcome! Stephanie Simms

Home My Dashboard DMP Requirements Public DMPs News Help Contact Us About Log Out

Woods Hole Oceanographic Institution

UC3 Contact UC3

My Dashboard | My DMPs | Create New DMP | My Profile

SELECT DMP TEMPLATE

Select one of the funder DMP Templates listed to proceed to the next step. The type of template chosen can affect what information you will need to provide in the following steps.

A - F G - L M - S T - Z All

- ☐ Alfred P. Sloan Foundation
- ☒ Department of Energy (DOE)
- ☐ Department of Transportation
- ☐ GoMRI Research Consortia DMP Template 2015
- ☐ Gordon and Betty Moore Foundation
- ☐ Institute of Education Sciences (US Dept of Education)
- ☒ Institute of Museum and Library Services
- ☒ National Institutes of Health
- ☒ National Science Foundation
- ☐ NEH-ODH: Office of Digital Humanities
- ☐ NOAA Data Sharing Template
- ☐ U.S. Geological Survey DMP Guidance
- ☒ University of California, Office of the President
- ☐ USDA - NIFA: National Institute of Food and Agriculture

<< Back Next >>

Select Funder

DMPTool Welcome! Stephanie Simms

Home My Dashboard DMP Requirements Public DMPs News Help Contact Us About Log Out

Woods Hole Oceanographic Institution UC3 Contact UC3

My Dashboard My DMPs Create New DMP My Profile

DMP Overview

DMP OVERVIEW

DMP Overview

DMP Template NSF-AGS: Atmospheric and Geospace Sciences

* DMP Title Experiments on atmospheric concentrations of CO₂

Proposal Solicitation Number

Proposal Submission Deadline

Visibility ☐ public ☐ institutional ☒ private

Add Co-owners: Add co-owners; search by name or email address

<< Back Cancel Save and Next >>

Fill out basic information about the plan

DMPTool Welcome! Stephanie Simms

Home My Dashboard DMP Requirements Public DMPs News Help Contact Us About Log Out

Woods Hole Oceanographic Institution UC3 Contact UC3

My Dashboard My DMPs Create New DMP My Profile

DMP Overview

DMP OVERVIEW

DMP Overview

DMP Template NSF-AGS: Atmospheric and Geospace Sciences

* DMP Title Experiments on atmospheric concentrations of CO₂

Proposal Solicitation Number

Proposal Submission Deadline

Visibility ☐ public ☐ institutional ☒ private

Add Co-owners: Add co-owners; search by name or email address

<< Back Cancel Save and Next >>

Co-owners can

- Edit
- Preview
- Download

Add co-owners

DMPTool Welcome! Stephanie Simms

Home My Dashboard DMP Requirements Public DMPs News Help Contact Us About Log Out

Woods Hole Oceanographic Institution UC3 Contact UC3

My Dashboard My DMPs Create New DMP My Profile

DMP Overview

DMP OVERVIEW

DMP Overview

DMP Template NSF-AGS: Atmospheric and Geospace Sciences

* DMP Title Experiments on atmospheric concentrations of CO₂

Proposal Solicitation Number

Proposal Submission Deadline

Visibility ☐ public ☐ institutional ☒ private

Add Co-owners: Add co-owners; search by name or email address

<< Back Cancel Save and Next >>

Three options:

1. Private – owners & co-owners only
2. Within institution
3. Publicly

Visibility settings

My Dashboard My DMPs Create New DMP My Profile

DMP Overview DMP Details DMP Preview

DMP DETAILS

NSF-AGS: Atmospheric and Geospace Sciences

Click on a section below to edit it at any time.

• = Complete
* = Mandatory

Template Outline

- ☒ Products of research
- ☐ Data format
- ☐ Access to data, and data sharing practices and policies
- ☐ Policies and provisions for re-use, re-distribution and production of derivatives
- ☐ Archiving of data

Instructions Links

Describe the types of data and products that will be generated in the research, the space and/or time-dependent information on chemical and physical processes, intermediate numerical results, theoretical formalisms, computational strategies, materials.

Guidance box size: small medium full

Describe what data or other research products you will generate in the course of your project. Include the size or amount of data produced, the type of data files that will be generated, and where and when the data will be produced. Examples of research products include observational data, results from models, data generated from previous observations or models, physical samples.

Answer questions

Cancel Changes Save Response Save and Next

<< Back DMP Preview >>

Instructions & help

My Dashboard | My DMPs | Create New DMP | My Profile

DMP Overview | DMP Details | **DMP Preview**

DMP PREVIEW

Export as PDF | Export as DOCX | Print

Export plan .pdf, .docx, .html

DMP Name: Copy of Making Data Count: Developing a Data Metrics Plan
 DMP Template: NSF-GEN: Generic
 Owner: Stephanie Simms
 Institution: University of California, Office of the President

Last Modified Date: April 21, 2016
 Co-owner(s):
 Solicitation Number: Early Concept Grants for Exploratory Research (EAGER)

Preview plan

used or exploited in the course of this project.

in the form of quantitative and qualitative survey results collected as part of Unit One. Given that these surveys will involve human subjects, we will ensure Institutional Review Board policies and receive prior approval before conducting the surveys. These data will be collected via paper and electronic means.

The second is dataset usage metrics harvested from the DataONE network through mechanisms developed in Unit Two. Individual DataONE member nodes collect raw web log usage statistics. Since this data potentially could be used to identify individual data consumers it is anonymized before being aggregated by the DataONE coordinating nodes that will be harvested by the DLM tool.

The third consists of software produced or modified as part of the project. The new DLM software created by PLOS will be maintained in a community GitHub repository during and after the project. Extensions to DataONE code will be maintained in the public DataONE Subversion repository. Finner will be responsible for code related to the DLM tool.

The fourth is dataset curation data harvested by the DLM tool from traditional and alternative (e.g., blogs, Twitter, etc.) channels of scholarly communication as part of Unit Three activities.

The fifth consists of all other research products (e.g., additional end-user tools developed in Unit Five, analysis results from Unit Five, community feedback, and project status communications) that will be made publicly available via either peer-reviewed journal articles or the project's website, which will be maintained by California by Simms.

Data and metadata standards

The DLM software produced by the project will conform to accepted community best practices, including version control (using git) with tagging of major releases, a permissive open-source license (Apache 2), public availability in a community repository (GitHub), in-line comments, reference and tutorial documentation with download and installation instructions that is available from within the software and from a community website, and extensive test coverage.

As part of this project the DataONE development team will evaluate the utility and level of effort necessary to report data usage statistics in a form that is consistent with COUNTER standards. This would permit meaningful comparisons between metrics for data usage and other types of online resources, including traditional serial and monographic publications.

Policies for access and sharing

All five major data categories, as enumerated in Section 1, will be publicly available for review, evaluation, and use as they are generated during the project and after its completion. Announcements about software and data availability will be using a variety of channels (e.g., blogs, Twitter, email lists) targeting all interested stakeholder communities.

As part of this project the DataONE development team will evaluate the utility and level of effort necessary to report data usage statistics in a form that is consistent with COUNTER standards. This would permit meaningful comparisons between metrics for data usage and other types of online resources, including traditional serial and monographic publications.

Policies for re-use, redistribution

Similarly, all software products resulting from this project will be re-usable and redistributable both during the project and after its completion. The only restriction placed on redistribution of the software is that the copyright and license statement be kept intact as required by the Apache open source license. This software is expected to be of interest to publishers, data centers and repositories, individual researchers, and institutional administrators.

As part of this project the DataONE development team will evaluate the utility and level of effort necessary to report data usage statistics in a form that is consistent with COUNTER standards. This would permit meaningful comparisons between metrics for data usage and other types of online resources, including traditional serial and monographic publications.

Plans for archiving and preservation

This DMP is **Private** | **Change Settings** | Note: when visibility is set to "Public", your DMP will appear on the Public DMPs page of this site and it will be downloadable and copy-able.

What else?

- Collaborative tool for co-PIs at different institutions or within a lab
- Keep DMPs from successful proposals for reference
- Customize content and custom templates
- Approval/review process with commenting system

blog.dmptool.org

DMPTool Blog

Guidance & resources for your data management plan

DMP Tool

Search

About the DMPTool

Go to the Tool

DMPTool Webinars

Presentations and Press

Maintenance

NIH Policy on Rigor and Reproducibility

Posted on June 2, 2016 by stephaniesimms

You've all heard about the reproducibility crisis in science. But you may not be aware of a (relatively) new National Institutes of Health (NIH) policy designed to address the issue. The NIH Policy on Rigor and Reproducibility became effective for proposals received on or after January 25, 2016 and applies to most NIH and Agency for Healthcare Research and Quality (AHRQ) grant applications. We just learned about the policy ourselves thanks to the combined efforts of UCSD library and research staff to raise awareness on their campus [and here's a noteworthy mention in a Nature review of 2015 science news]. To aid researchers in meeting the new criteria, UCSD produced this [handy guide](#) that we (and they) would like to share with the wider community.

The new policy does not involve any changes to data sharing plans. It is related and important enough, however, that we inserted a statement and link in the "NIH-GEN: Generic" template (Please note the Rigor and Reproducibility requirements that involve updates to grant application instructions and review criteria [but not Data Sharing Plans]).

The policy does involve:

- Revisions to application guide instructions for preparing your research strategy attachment
- Use of a new "Authentication of Key Biological and/or Chemical Resources" attachment (example from UCSD library website)
- Additional rigor and transparency questions reviewers will be asked to consider when reviewing applications

These policies are all meant to achieve basically the same goals: to promote openness, transparency, reproducibility, access to, and reuse of the results of scientific research. We're grateful to the folks at UCSD—Dr. Anita Bandrowski, Ho Jung Yoo, and Reid Otsuji—for helping to consolidate the message and for providing some new educational resources.

Posted in [News](#), [OSTP/Resp](#) | Tagged [communication](#), [funder update](#), [grant update](#), [resources](#), [website](#) | [Leave a reply](#)

Email | uc3@ucop.edu
 Twitter | [@TheDMPTool](https://twitter.com/TheDMPTool)
 Blog | blog.dmptool.org

DMPTool.org
 My thoughts: goo.gl/SLi2jP

Presentation Adapted from:
<https://dmptool.org/promote#slides>